DOCUMENT RESUME

ED 203 928 JC 810 371

AUTHOR Watts, Gordon E.; Hammons, James D.
TITLE Staff Development: A Time for Appraisal.
INSTITUTION American Association of Community and Junior Colleges, Washington, D.C. National Council for Staff, Program and Organizational Development.
PUB DATE 16 Nov 80
NOTE 30p.; Report prepared for the National Conference of the National Council for Staff, Program and Organizational Development (Des Moines, IA, November 16, 1980).
AVAILABLE FROM Vice President for Publications, NCSPOD, Hocking Technical College, Nelsonville, OH 45764 ($2.00).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Evaluation Criteria: Evaluation Methods; Evaluation Needs: Literature Reviews: Models; Program Effectiveness; *Program Evaluation; *Staff Development; *Systems Approach: Two Year Colleges

ABSTRACT

After noting the lack of attention paid to the evaluation of staff development programs in practice and in the literature, this monograph presents a conceptual framework for planning program evaluations. The monograph first discusses the purposes of staff development evaluation and then considers three prerequisites for the development of an evaluation plan: institutional goals and objectives, staff needs assessment, and a plan delineating program objectives and guidelines. The proposed systems model for program evaluation is then presented, in which the evaluator measures both the effect of staff development on participants and the impact of acquired skills on the work setting. The monograph then delineates guidelines and problems in gathering data at four levels of evaluation: participant reaction to activities, amount of participant learning, changes in behavior through application of this learning, and results. A matrix is then presented, detailing the questions to be asked in gathering and assessing information at each level. The monograph concludes with a discussion of summative vs. formative evaluation, securing needed resources, the politics of evaluation, and the extent and timing of staff development evaluation activities. A staff development program evaluation form, a pre-/post-test for self-assessment of learning, and an annotated bibliography are included. (JP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203928

STAFF DEVELOPMENT: A TIME FOR APPRAISAL

By
Dr. Gordon E. Watts
Westark Community College
Fayetteville, Arkansas

Dr. James D. Hammons
University of Arkansas
Fayetteville, Arkansas

Nov. 16, 1980
NCSPOD National Conference
Des Moines, Iowa

NCSPOD is an Affiliate Council of
the American Asscciation of Community and Junior Colleges

JC 810 371

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Candace S. Vancko

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

STAFF DEVELOPMENT: A TIME FOR APPRAISAL

A 1975 ETS survey by John Centra reported that close to 60 percent of 2600 degree granting institutions of higher education had some type of staff development program or someone who coordinated staff development (Centra, 1976). As the decade of the 80's begins, there are no indications that staff development has experienced the loss of momentum so characteristic of other innovative ventures in higher education; rather, the interest in the development of staff is continuing to grow. Evidence of this trend include the formation of two national organizations, the Professional and Organizational Development Network in Higher Education and the more recently organized National Council for Staff, Program and Organizational Development of the American Association of Community and Junior Colleges; the exponential increase in the number of publications and conferences on the topic; and the establishment of two university directed institutes for professional development, the National Institute for Staff and Organizational Development at the University of Texas and Memphis State University's Institute for Academic Improvement.

A partial explanation for the current acceptance of staff development is its underlying assumption that improvements in the professional and personal lives of staff will lead to more effective and efficient operation of the institutions in which they work. Past deficiencies in pre-service preparation and gross neglect of in-service education, coupled with the pressures of a "steady state" environment and new demands for accountability, have also contributed to providing an unusually receptive environment for staff development by trustees, administrators, and faculty.

However, for newly established staff development programs to persist beyond their present stage, they must prove their worth--that is, they must be evaluated. Colleges need to know how effective staff development programs are, what impact the programs have on participants as well as the institution, whether or not an acceptable ratio of program costs to program benefits exists, and ultimately, what measurable benefits accrue to students, such as improved learning, improved employability, or increased retention, although the latter is an institutional benefit as well.

To date, the response to the need for evaluation has been disappointing. Evidence of this point is provided by Centra (1976) who, in an assessment of the practices of faculty development programs in 756 institutions of higher education, found that only 14% had completed evaluations of their programs or activities, while another 33% of the institutions had completed partial evaluations. Thus, slightly more than half of the faculty development programs had no evaluative component at all.

The reasons for this lack of evaluation are several. A major factor is, as Gaff (1976) points out "... promoters of institutional improvement programs have been too busy getting things in motion to worry about evaluating what they are doing." Other contributing factors include the non-evaluation orientation of many of the campus and national leaders of staff development, and the lack of a theoretical or practical literature foundation on which to base an evaluation. The latter is perhaps the major causal factor for the present lack of evaluation data. In the scores of higher education articles, monographs and books on staff and faculty development, only two provide more than a cursory look at program evaluation (Hammons, Wallace and Watts, 1978; Smith, 1977).

This article, written in an effort to stimulate further attention to the evaluation of staff development programs, represents an attempt to provide both an initial conceptual base for the evaluation of staff development and a guide for the busy practitioner. After a brief discussion of the purposes and prerequisites for evaluation, we present a proposed model for evaluating staff development programs. This is followed by a discussion of several considerations to be observed in evaluating a program. Finally, we provide an annotated bibliography of carefully selected sources on evaluation.

PURPOSES FOR EVALUATION

Newstrom (1978), in an article written for staff training in a business and industry setting, suggested eight different reasons for evaluating training programs. A number of these (with necessary adaptations) seem equally relevant to higher education.

1. To assess the achievement of program objectives.
2. To assess the effectiveness of the staff development facilitator or director.
3. To justify the expense of staff development through a cost-benefit analysis.
4. To improve the content or structure of the staff development program.
5. To decide if other staff members could or should participate in the program.
6. To identify who benefited the most or least from the program.

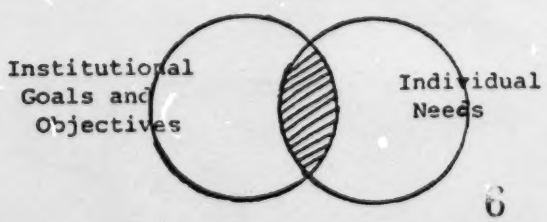
Most of the above reasons can be subsumed under the heading of formative and summative evaluation, terms first coined by Michael Scriven (1967) to indicate two distinct types of evaluation. Formative evaluation takes

place continuously during the development and implementation of a program. Its purpose is to provide decision making information in order to make improvements or adjustments in the program's plans, activities, and/or anticipated outcomes. Summative evaluation takes place after a program has been completed. Its purpose is to secure data necessary to determine whether a program as implemented to date should be continued, terminated, replicated, or perhaps disseminated. Both purposes should be accommodated in evaluating staff development programs.

PREREQUISITES FOR EVALUATION

In order to develop a valid evaluation plan, three prerequisites are needed -- institutional goals and objectives, a valid needs assessment of the staff, and a staff development plan containing a statement of philosophy, the objectives of the program, the scope of the program, and any necessary guidelines for program implementation. These are prerequisites because they are necessary in order to determine what is to be evaluated as well as to establish the criteria for success.

Briefly, the way in which these prerequisites interact with one another to facilitate evaluation is as follows: Every organization has certain goals and objectives, and individuals within that organization have certain job-related needs. When those institutional goals and objectives are clearly stated and made known, and when individual needs are determined through a needs assessment, then a staff development plan can be derived. In developing the plan, the focus should be on the intersection of institutional goals and individual needs as shown below:



From this intersection would come the goals and objectives of the staff development program. If developed properly, each goal and objective would be written to reflect the activity involved, the anticipated results of that activity and the standards used to judge the success of that activity. For example: as a result of a workshop on motivational techniques in the classroom, at least 10 instructors will implement the techniques and within two semesters will have increased student retention in their classes by 7%. Knowing, therefore, what to evaluate and the criteria to use, the essential groundwork for evaluation would be established.

A valid evaluation plan should evolve after fulfilling the above prerequisites. Ideally, the evaluation should be systematic, with fairly well defined steps, procedures, or guidelines for the evaluator to follow. The term most often used in the literature on evaluation to describe such frameworks or guidelines is "model." In looking for models, we first turned to the literature of higher education. Smith (1977) describes three evaluation models that can be used. They are: formative and summative evaluation, discussed earlier in this article; goal-free evaluation, which focuses attention not only on the intended goals of a program but also those unintended, unplanned for outcomes; and the medical model, a holistic approach that analyzes what goes into a program, its outcomes, and all other factors which would influence the success of the program. However, each of these, while containing useful ideas, is insufficient in itself to satisfy the requirements of staff development program evaluation.

Another source of possible models was the literature on evaluation. Included here would be the Countenance Model (Stake, 1967), the Discrepancy Model (Provus, 1971), the Goal Attainment Model (Glaser, 1970), the CIPP Model (Stufflebeam, 1971), and the decision-oriented model developed by UCLA's Center for the Study of Evaluation. Each of these models, although using different terminology and varying in their complexity, contains several common requirements; i.e., stating specific evaluative criteria, identifying conditions that exist prior to program implementation, evaluating what happens during the program, and analyzing outcomes upon program completion. Any of these models could be adapted for use in evaluating a staff development program. However, to utilize the models properly may require more time and more knowledge of evaluation than most personnel in staff development presently possess.

Having reviewed the literature of higher education and general evaluation and found it wanting, we then turned to training literature from business and industry and found an abundance of relevant sources. On reflection, this should have been expected, since business and industry have had staff development programs for a long time, and profit minded, effectiveness-efficiency-oriented managers have been demanding evaluation of staff development activities and budgets for years. We found the work of Kirkpatrick (1967) and Brethower and Rummler (1977) to be particularly helpful. Kirkpatrick was the first to conceptualize different levels of evaluation, while Brethower and Rummler's work combined some of Kirkpatrick's ideas and systems theory into an evaluation matrix. Their work forms the basis of the evaluation model which follows.

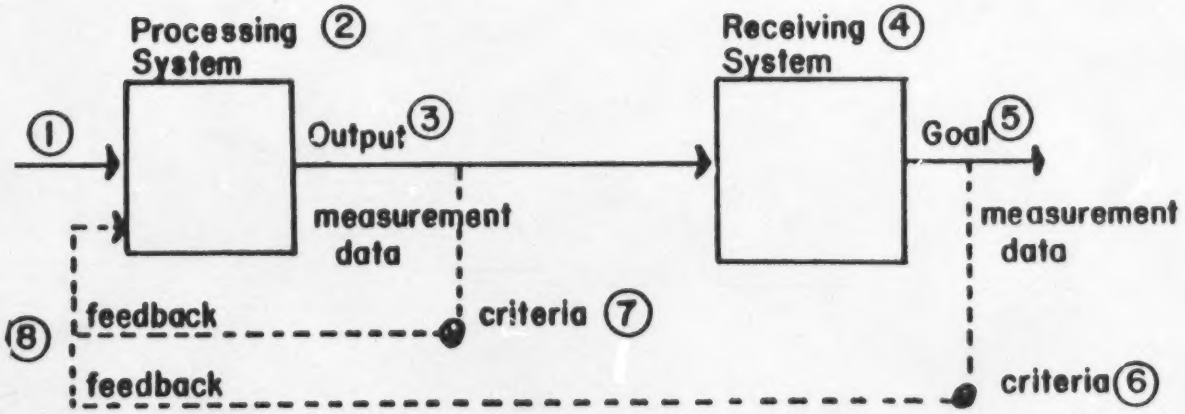
A SYSTEMS MODEL FOR EVALUATING STAFF DEVELOPMENT PROGRAMS

Brethower and Rummler (1977) point out that much of the confusion surrounding the evaluation of staff development is because "... people can't agree on what they are trying to evaluate and why" and consequently they "... won't agree on how to evaluate." They state, however, that if staff development is viewed in terms of general systems theory (input, process, output), a number of alternatives for evaluating are available.

Figure 1 illustrates, in using a systems framework, the ideal relationship between staff development and the institution it serves. The key components of the system are the processing system which is the staff development program and the receiving system, which consists of the specific jobs within the institution. In some instances, such as organizational development, the institution itself serves as the receiving system. Specifically, the components as described by Brethower and Rummler are:

1. Inputs into the system, such as instructors, administrators, or secretaries.
2. The processing system, which converts inputs into outputs through such means as workshops, conferences, or seminars.
3. The outputs of the processing system, which are those same instructors, administrators, or secretaries with newly acquired skills, behavior, or knowledge.
4. The receiving system, which is the work setting into which the outputs go. (It is important to note that the processing system and the receiving system are actually subsystems of a larger system which in most cases is the institution).

5. The stated goal of the receiving system, such as "student dropouts will decrease by 10%," where the receiving system is the classroom and the processing system is an instructional workshop.
6. The evaluation of the stated goal of the receiving system (e.g., do the dropouts actually decrease by the expected 10%?).
7. The evaluation of the outputs of the processing system. (The assessment here would focus on whether or not or to what degree the participants achieved what they were supposed to as a result of the workshop).
8. The feedback to the processing system regarding the outputs of both the receiving system and the processing system.



Brethower and Rummler advocate the systems viewpoint discussed above for several reasons. First, considering the outputs of staff development as inputs of another system emphasizes the fact that staff development cannot and does not function in a vacuum. It must function as an integral part of a larger system which is the institution.

Also, without the receiving system as a part of the model, there is no way to determine the value of the staff development program as a processing system to the institution. For example, an evaluation plan without the receiving system concept would only allow for evaluation of the immediate outputs of staff development, such as mastery of the program objectives, or might stress the volume of staff development activity and/or its popularity, thus failing to consider the impact of the program upon the needs of the institution.

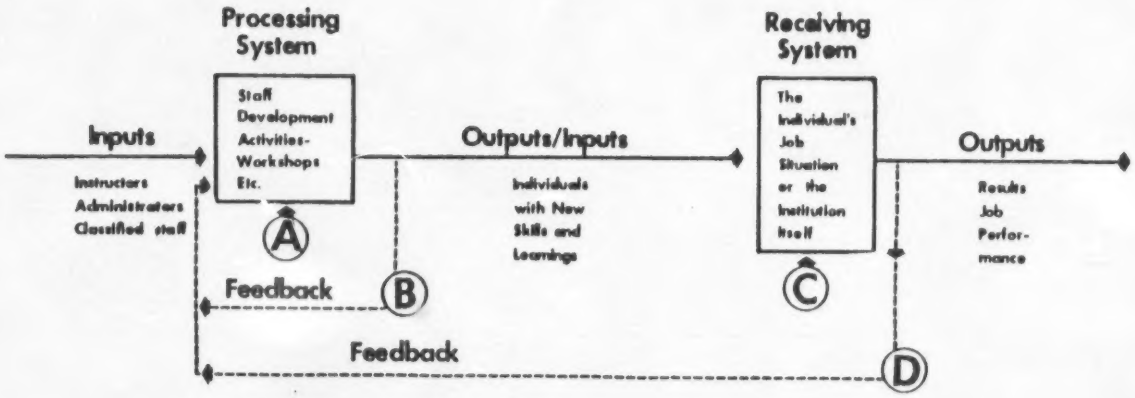


FIGURE 2
(Adapted from Brethower and Rummler, 1977)

Finally, the representation of staff development as a system, as shown in Figure 2, reveals several sources from which to gather evaluative data--the processing system and its outputs, and the receiving system and its outputs. These sources, labeled A through D, are identical to the four levels of evaluation (reaction, learning, behavior and results) originally described by Kirkpatrick (1967). Each level has distinct criteria for evaluating staff development and can furnish data for either formative or summative purposes. The following discussion explains each of the four levels in some detail.

LEVELS OF EVALUATION

Level A - Reaction

The reaction level assesses how the participants in staff development activities feel about those activities. An evaluation of a workshop on cognitive mapping, for example, would assess participants' feelings regarding such factors as the enthusiasm of the workshop leader, the use of visual aids, the clarity of the workshop objectives, the amount of material covered, and so forth.

CHARACTERISTICS OF THE FOUR EVALUATION CRITERIA

Criteria	Value of Information	Frequency Of Use	Difficulty Of Assessment
A. Reaction	Lowest	Relatively Frequent	Relatively Easy
B. Learning	↓	↑	↓
C. Behavior	↓	↑	↓
D. Results	Highest	Relatively Infrequent	Relatively Difficult

FIGURE 3
(Newstrom, 1978)

As shown by Figure 3, evaluation at the reaction level is the easiest to conduct compared to the other levels. It also yields the lowest informational value and is the most frequently used. But its frequency of use does not guarantee that it is done properly. Kirkpatrick (1967) suggests the following five guidelines for evaluating the reaction level:

1. Determine what facets of the activity you want to assess.
2. Develop a form to assess them.
3. Design the form so that reactions can be tabulated and quantified.
4. Maintain the anonymity of the participants for more honest reactions.
5. Allow the participants to write additional comments not covered by the other portions of the form.

Since the above guidelines advocate the use of some type of questionnaire, an example of what a typical form might look like is shown in Figure 4.

Level B - Learning

Once data have been gathered at the reaction level, the evaluator has information regarding how well the program was received as well as information that can help improve the program. However, a positive reaction to the program does not necessarily mean that the participants learned anything.

To extend the example of the cognitive mapping workshop. The participants may have thoroughly enjoyed the workshop because it had multiple visual aids, numerous handouts, and a leader that commanded everyone's attention. But, they may not have learned anything about cognitive mapping. So, the central concern here is whether or not the participants learned what the program indicated they were supposed to have learned. Again, Kirkpatrick (1967) offers a set of guidelines for measuring learning:

1. Measure the learning of each participant so that quantitative results can be determined.
2. Utilize a pre-and-post approach to relate learning to the activity or program.
3. Measure the learning on an objective basis as much as possible.
4. Utilize a control group when possible for comparison to the group that participated in the activity.
5. Analyze the results statistically so that results have more credibility.

Although a number of methods could be used to assess learning, a simple form, such as that shown in Figure 5, given before and after a particular staff development program, can be used to provide adequate information regarding participants' opinions about their learning. Obviously, to actually test the aquisition of knowledge and skills, or to measure attitudinal changes, more sophisticated pre-and-post testing would be required.

Level C - Behavior

There is a great difference between learning a new skill or gaining new knowledge and putting that skill or knowledge to use. Therefore, the next logical place from which to gather evaluative data is the job setting. The focal point of evaluation at the behavior level, then, is whether or not participants change their behavior (as a result of having gained knowledge or skills through a staff development program). For example, do participants apply cognitive mapping concepts or techniques learned from the workshop to their classroom practices?

Evaluation at this level starts yielding more useful information about a staff development program, but at the same time becomes more difficult to implement. The guidelines that Kirkpatrick (1967) outlines for assessing

behavior changes are:

1. Job performance should be appraised both before and after the staff development program.
2. Job performance should be appraised by a number of people familiar with the participant's job.
3. Before and after job performance should be statistically analyzed in order to relate it to the staff development program.
4. Appraisal of job performance should take place long enough after the program for any changes to take place.
5. A control group who does not participate in the program should be used.

The difficulty in evaluation at this level is made more complex due to the need for data acquired primarily from observation -- and the inherent measurement problems associated with that.

Level D - Results

Assessing the results of a staff development program is the most difficult to accomplish. And, compared to the other levels, the results level yields the most valuable information. Unfortunately, results are not frequently determined, partly because of a lack of prerequisite goals and objectives. As stated previously in the discussion of prerequisites, staff development objectives should state the results that are anticipated. When the objectives are stated in specific terms, such as increased retention, increased student learning or improved cost effectiveness, evaluation at the results level becomes easier.

But, the main concern, determining what has happened as a result of applying concepts learned through an enjoyable staff development program, is still difficult. For example, what are the effects that the application of cognitive mapping concepts have on an instructor's job performance? Has instruction improved? Have students learned more? Have dropouts decreased? If so, how do we know?

Unlike evaluation at the other levels, Kirkpatrick offers no specific guidelines to follow in assessing results. He suggests that if the results or criteria have been previously stated, then evaluation should be similar to that at the behavior level. Otherwise, Kirkpatrick feels that the difficulty of separating variables and trying to specify which results can be attributed to which variables renders the task almost impossible.

Ideally, evaluation should include data from the reaction level to the results level. In reality, most evaluation efforts never go beyond the reaction level. There are a variety of reasons for this neglect, but one of particular relevance to our discussion was mentioned in a recent article by John Newstrom (1978). He states that many staff development practitioners make the assumption "... that there is a high sequential intercorrelation..." among the evaluation levels. Such an assumption leads to the following thought process: If the reaction to a staff development activity is favorable, then participants will probably learn more; if they learn more, then their behavior will probably change; and if their behavior changes, the anticipated results will follow. Naturally, a reverse set of conclusions would result from an unfavorable reaction.

Newstrom is quick to point out that there is a "Catch-22" inherent in such an assumption and thought process. According to him, the ultimate purpose for evaluating staff development programs is to reach a conclusion regarding the effectiveness of the program. If evaluative data are obtained at all four levels, and if a simple conclusion is reached at each level (e.g., favorable or unfavorable reaction, increased learning or no learning, desirable change in behavior or no change, or improved results or no improvement), then sixteen different possibilities exist for a set of evaluative results.

For example, reaction could be unfavorable, learning increased, behavior changed, but no change in results. How should a program with that type of evaluative data be assessed in terms of effectiveness? If only reaction was assessed, the program could have been deemed unsuccessful. However, learning did occur. If only learning and behavior were assessed, the program may have been praised. But, the program achieved no results. The critical point here is to be cautious when utilizing results from only some levels to reach a conclusion regarding program effectiveness. Assumptions made on incomplete data could be erroneous. The preferred approach, therefore, is to make the collection of data at all levels not just the ideal, but rather, the standard.

The Evaluation Matrix

Once the levels of evaluation have been established to form the basis for the evaluation plan, an evaluation matrix can be designed by asking the following questions at each of the four levels (Brethower and Rummler, 1977).

1. What do you want to know? This is the basic question asked at each evaluation level. For example, the basic question at the reaction level is: do participants like the staff development activity?
2. What can be measured to answer those questions?
3. What dimensions of learning or performance are to be measured?
4. What are the sources of the measurement data?
5. What ways are the data to be gathered?
6. What evaluation standards are to be applied to each question?

The resulting matrix which operationalizes the model is shown in Figure 6. For illustrative purposes, the matrix depicts how the total

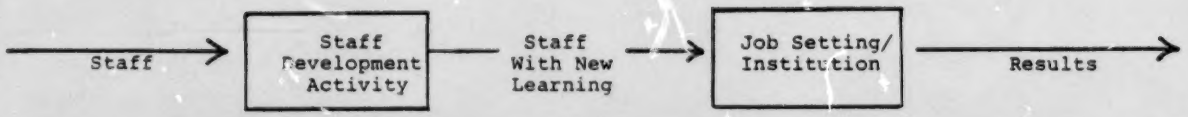
model could be applied in devising an evaluation for a cognitive mapping workshop. Thus, the matrix serves as a guide to determine precisely what and how to evaluate staff development.

Up to this point, we have suggested that a systems model of staff development which includes four different levels for evaluation should be followed. We have further suggested that by utilizing a matrix, the crucial specifics of what and how to evaluate can be determined. The remainder of this article focuses on some considerations that need to be taken into account as the model is utilized to plan staff development evaluation.

CONSIDERATIONS

PURPOSES OF EVALUATION

Without a clear statement of purpose to give direction, evaluation efforts are likely to be focused toward gathering the wrong data or using data incorrectly, thus rendering the results of little value. The three major purposes for evaluation are judgmental, developmental and informational. Judgmental and developmental evaluation are synonymous to summative and formative evaluation; however, informational evaluations have a primary purpose of collecting normative data for informing others of the status of the program rather than to improve the program or to decide on its continuation or amount of future funding.



Level	Reaction	Learning	Behavior Change	Results
What do you want to know?	Do participants like the activity or program? If not, why not?	Did learning occur? If not, why not?	Are the learnings applied in job setting? If not, why not?	Does application of learning have any measurable effects? If not, why not?
What can be measured?	Participant reaction during or after the workshop	Participant's knowledge or performance after the workshop as contrasted with before the workshop	Extent of mapping usage in classroom after training as contrasted with before training	Student performance Student attrition
Dimensions of measurement	Relevance of content Workshop design Competence of resource persons	Understanding of theory Application of principles	Attempts to match learning style with instructional style Provisions made available for alternate learning pathways	Grades Course completion time % of material learned Student reaction
Sources of data	Responses to workshop evaluation forms Comments to other participants Comments to resource persons Attendance Attentiveness	Performance on exercises Presentaitons to other participants Post test results	Classroom behaviors and methodologies utilized Instructional materials developed	Student records Instructor records
Data gathering methodology	Observation Interview Questionnaires	Observation Document review Questionnaires Objective Test	Observation Interview Review of instructional materials produced	Interview Statistical compilation of data from student/instructor records
Evaluation criteria	At least 80% of participants should respond favorably	At least 90% of participants will demonstrate learning of 90% of content presented	At least 50% of participants will utilize cognitive mapping in at least one class within 18 months	No. of students achieving A.B.C.'s will increase 15% Attrition will drop 10% in a year

FIGURE 6
(Adapted from Brethower and Rummler, 1977)

Stage of Program Development

A staff development program that is in its infancy should place heavy emphasis on formative evaluation. At this stage, the program leaders are trying to maintain a favorable image, offer useful activities, and, in general, improve the program as much as is possible or necessary. For those purposes, evaluation at the reaction and learning levels would be more beneficial.

As the program matures, the emphasis should shift to summative evaluation. Program leaders at this point should be determining whether or not participants are applying what they have learned and whether certain activities should be continued. Thus, evaluation at the behavior and results levels should take precedence for two reasons: the upper administration is at this stage wanting to know if the program is worth continuing, and the program leaders are seeking to make determinations regarding the continuance of specific staff development activities.

Resources

Conducting an evaluation, whether of staff development or any other educational endeavor, requires the use of certain resources; namely time, money, and expertise. Time is necessary to plan, execute, and analyze the results of an evaluation. Naturally, since the behavior and results levels of evaluation are more difficult to utilize, more time is required for them than if only the reaction and learning levels are used.

Financing of evaluation is also important. As with time, the higher levels of evaluation will require more expenditures, especially if external evaluators are used. It is quite possible that designing and implementing a thorough evaluation could require so much time and financial resources that it would have a detrimental effect on the program.

For those not skilled in evaluative techniques, a resource person to assist in developing the evaluation plan can be an invaluable asset. Personnel from an office of institutional research can often provide the most assistance, since they are usually well grounded in research and evaluation. However, if an institutional researcher is not available, then a faculty or staff member well versed in research techniques might be called upon for assistance. If those avenues yield no one to assist, then a consultant from outside the institution could be utilized.

One further resource worthy of mention are those who practice staff development. As evaluation becomes more prevalent, staff development personnel can share forms, techniques, and expertise with one another. (Toward that end, a carefully selected annotated resource bibliography is included at the end of this article to help make the job of evaluation more clear and understandable.)

Politics of Evaluation

Since staff development does not take place in a vacuum, it is necessary to consider the political aspects of evaluation. One consideration is determining who will actually conduct the evaluation. It is often assumed that the person responsible for the staff development program would also evaluate it. However, others may be involved, such as outside consultants, faculty members, a Staff Development Committee, or some other group. If this is the case, then the staff developer must be certain that that person or group has a clear understanding of the purpose of evaluation (formative or summative) and that they can be trusted to maintain an objective viewpoint throughout the evaluation process. Otherwise, an evaluation may be so biased as to be meaningless.

Another political consideration is the nurturing of continued support from the administration. The staff developer should assume that at least some administrators will want to see data regarding program effectiveness whether they ask for it or not. Providing such data on a regular basis will not only prevent the staff developer from being put in the embarrassing position of having no data on hand when an administrator requests it, but also can counter some negative attitudes toward the program.

Any evaluation reports prepared for the administration should be clear and concise so that the administrator can learn in a short space the essence of the evaluation. Finally, the staff developer should be sure that the reports are sent to any administrator who could have a potentially influencial voice in any decision regarding program continuance.

Extent of Evaluation

Another consideration is whether it is necessary to provide evidence of program effectiveness or proof of it. Kirkpatrick (1977) points out the distinctions between the two. Evidence is data that supports the notion that participants liked a staff development activity, learned the material presented, and applied it on the job with positive results. Evidence is easy to gather; for example, simply asking program participants if they changed their behavior after attending a staff development activity will provide evidence of any changes. Proof, however, requires more. To obtain proof that a program produced changes in behavior, the evaluator must obtain a measurement of the behavior both before and after the program. Then, and this is the crucial aspect of proof at every level, it is necessary to indicate that the specific staff development program or activity in question and no other possible alternative is responsible for the changes in behavior.

Naturally, gathering proof will require more stringent evaluative procedures and will be more time consuming and expensive than gathering evidence. Kirkpatrick suggests that proof should be the ultimate goal, but recognizing that it is sometimes impossible or impractical, evidence is satisfactory.

Timing

A prime consideration in evaluating staff development activities is when to conduct the evaluation. For the reaction and learning levels, evaluation should take place as soon after the activity as possible. Evaluation at the behavior and results levels will clearly involve longer time periods, for time must be allowed for behavior to change and results to become manifest.

CONCLUSION

Staff development personnel are in the position to make a significant contribution to education in the decade of the 80's. Staff development has the potential to facilitate program, instructional, and organizational development. But, these personnel must evaluate the effectiveness of their programs in order to "... establish the worth and credibility of staff development" (Case, 1978). A necessary first step is to take evaluation beyond the reaction level where most programs, if they evaluate at all, seem to be. Evaluation must be extended to the higher levels so that program effectiveness can be determined.

The model presented above should provide the basis for a sound evaluation plan. If the model is used and tempered with a proper balance of time, money and expertise, then staff development should be able to achieve its potential. Without such evaluation, staff development could become another short lived educational fad.

STAFF DEVELOPMENT PROGRAM EVALUATION FORM

Program: Date:

Leader:

We would like your opinion of certain aspects of the program represented by the statements below. Circle the number that best expresses your opinion according to the following scale.

1 = Strongly Disagree
2 = Disagree
3 = Agree
4 = Strongly Agree
5 = No Opinion/Not Applicable

1.	The purposes of the program were clear.	1	2	3	4	5
2.	The program objectives were relevant to me.	1	2	3	4	5
3.	The content of the presentations was consistent with stated program objectives.	1	2	3	4	5
4.	The general format of the program should be maintained.	1	2	3	4	5
5.	The length of the program was about right.	1	2	3	4	5
6.	The program provided a good balance of theory and practice.	1	2	3	4	5
7.	The group exercises contributed to the program's effectiveness.	1	2	3	4	5
8.	The handouts were helpful.	1	2	3	4	5
9.	The program leaders were responsive to the needs of the group.	1	2	3	4	5
10.	Leader presentations were effective.	1	2	3	4	5
11.	The leaders used audio-visual aids effectively.	1	2	3	4	5
12.	The program leaders seemed to be well-trained, knowledgeable, and generally competent.	1	2	3	4	5
13.	I would recommend this program to my colleagues.	1	2	3	4	5

14. Further comments or suggestions.

FIGURE 4

PRE- POST LEARNING ASSESSMENT

TOPIC: Increasing Student Motivation

LEADER:

Please rate your ability to perform the tasks indicated below by using the scale following each objective.

1. Describe Maslow's theory of motivation.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

2. Describe Herzberg's theory of motivation.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

3. Discuss the implications of Maslow and Herzberg's theories in working with students.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

4. Describe 10 classroom behaviors or instructional practices that can help to motivate students.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

5. Describe 10 classroom behaviors or instructional practices that demotivate students.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

6. Describe the characteristics of students who are success identifiers.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

7. Describe the characteristics of students who are failure identifiers.

   NOT AT ALL 1 2 3 4 5 6 7 8 9 10 VERY WELL

FIGURE 5

ANNOTATED BIBLIOGRAPHY

The following resources were carefully selected for their content and informational value in designing and implementing the evaluation of staff development programs and/or activities. All references to training and training programs can be read as staff development and staff development programs.

Brethower, Karen S. and Rummler, Geary A. "Evaluating Training." Improving Human Performance Quarterly, 1977, S, 3-4, 103-120.

A model is presented for identifying appropriate methods for evaluating training programs based on a systems approach of inputs, a processing system, outputs, and a receiving system. The model includes four levels of evaluation along with an evaluation matrix. A discussion of how to solve some common evaluation problems completes the article.

Bunker, Kerry A. and Cohen, Stephen L. "Evaluating Organizational Training Efforts: Is Ignorance Really Bliss?" Training and Development Journal, Vol. 32, No. 8, August, 1978, 4-11.

Several critical inadequacies in the evaluation of training are discussed with special emphasis on internal validity. The authors demonstrate how the inadequacies can be overcome and how evaluation designs can be improved through the use of better controls.

Kirkpatrick, Donald L. "Evaluation of Training" in Robert L. Craig and Lester R. Bittel (Eds.) Training and Development Handbook. New York: McGraw-Hill, 1967.

Kirkpatrick, Donald L. "Evaluation of Training" in Robert L. Craig (Ed.) Training and Development: A Guide to Human Resources Development. New York: McGraw-Hill, 1976.

Kirkpatrick's explanation of his evaluation model was first published in 1959-60 in the Training and Development Journal. It has since appeared as chapters in the above two books. In these chapters, Kirkpatrick thoroughly explains the reaction, learning, behavior, and results model, offers examples of forms that can be used for evaluation, and presents numerous examples of evaluation studies conducted in the business/industrial setting.

Kirkpatrick, Donald L. "Evaluating Training Programs: Evidence vs. Proof." Training and Development Journal, Vol. 31, No. 11, November 1977, 9-12.

Kirkpatrick discusses the difference between gathering evidence and proof at the reaction, learning, behavior, and results levels of evaluation. Although evidence is easier to gather, and proof may be impossible, the article describes ways of actually obtaining proof of training effectiveness.

Kirkpatrick, Donald L. (Ed.) Evaluating Training Programs. Madison, Wisconsin: American Society for Training and Development, Inc., 1975.

This highly informative compilation of articles provides a wealth of ideas and approaches for evaluating training. The book offers general articles on evaluation as well as full chapters devoted to descriptions of actual evaluations conducted at the reaction, learning, and behavior levels of Kirkpatrick's evaluation model.

Newstrom, John W. "Catch 22: The Problem of Incomplete Evaluation of Training." Training and Development Journal, Vol. 32, No. 11, November, 1978, 22-24.

The set of assumptions often made when utilizing Kirkpatrick's evaluation model can lead to a simplistic view of evaluation. The author clearly demonstrates the dangers of an incomplete or too simplistic view of evaluation. Arguing for more rigorous evaluation design, he offers suggestions for overcoming those dangers.

Rose, Clare and Nyre, Glenn F. The Practice of Evaluation. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurements, and Evaluation, 1977.

For those unfamiliar with educational evaluation, this volume offers one of the most concise introductions to evaluation, evaluation models, and evaluation designs available. The monograph is further strengthened by case studies of educational evaluations which take the theoretical aspects of evaluation and show how they have been applied in real situations.

REFERENCES

1. Brethower, Karen S. and Rummler, Geary A. "Evaluating Training." Improving Human Performance Quarterly, 1977, S, 3-4, 103-120.

2. Case, Chester H. "Staff Development for the Community College." In Jack Lindquist (Ed.) Designing Teaching Improvement Programs. Berkeley, California: Pacific Soundings Press, 1978.

3. Centra, John A. Faculty Development Practices in U.S. Colleges and Universities. Princeton, New Jersey: Education Testing Service, 1976.

4. Gaff, Jerry G. Toward Faculty Renewal. San Francisco: Jossey-Bass, Inc., 1975.

5. Glaser, Robert. "Evaluation of Instruction and Changing Educational Models." In M. C. Wittrock and D. E. Wiley (Eds.) The Evaluation of Instruction: Issues and Problems. New York: Holt, Rinehart, and Winston, 1970.

6. Hammons, James O., Wallace, Terry H. Smith, and Watts, Gordon E. Staff Development in the Community College: A Handbook. Los Angeles: ERIC Clearinghouse for Junior Colleges, 1978.

7. Kirkpatrick, Donald L. "Evaluation of Training." In Robert L. Craig and Lester R. Bittel (Eds.) Training and Development Handbook. New York: McGraw-Hill, 1967.

8. Kirkpatrick, Donald L. "Evaluating Training Programs: Evidence vs. Proof." Training and Development Journal. November, 1977, 9-12.

9. Newstrom, John W. "Catch 22: The Problems of Incomplete Evaluation of Training." Training and Development Journal. Vol. 32, No. 11, November, 1978, 22-24.

10. Provus, Malcolm. Discrepancy Evaluation for Educational Program Improvement and Assessment. Berkeley, California: McCutchan, 1971.

11. Scriven, Michael. "The Methodology of Evaluation." In R. E. Stake (Ed.) Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

JUL 24 1981

UNIVERSITY OF CALIFORNIA
ERIC CLEARINGHOUSE FOR
JUNIOR COLLEGES
96 POWELL LIBRARY BUILDING
LOS ANGELES, CALIFORNIA 90024

12. Smith, Albert B. "Evaluating Staff Development Programs." New Directions for Community Colleges, No. 19, Autumn, 1977.

13. Stake, Robert E. "The Countenance of Educational Evaluation." Teacher's College Record, 1967, 68, 523-540.

14. Stufflebeam, Daniel L., et al. Educational Evaluation and Decision Making. Itasca, Illinois: F. E. Peacock, 1971.